Research Article

Application of PCA-SVM Model in Financial Crisis Early Warning System of Listed Manufacturing Companies

Jing Li [ID][1,2] **and Mingliang Zhang** [ID][3]

[1]*School of Finance, Xi'an Eurasia University, Xi'an 710065, China*
[2]*Equipment Management and Unmanned Aerial Vehicle Engineering School, Air Force Engineering University, Xi'an 710051, China*
[3]*Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China*

Correspondence should be addressed to Mingliang Zhang; zhangml@jou.edu.cn

Manufacturing industry has always occupied this very important proportion in the national economy. However, with the economic problems in the past two years, some enterprises began to have a deficit crisis. In order to fundamentally alleviate and solve the financial problems of enterprises, this paper studies the financial model prediction of enterprises based on principal component analysis and support vector machine. The multiple financial analysis trajectories are constructed by using PCA-SVM model, and the results are compared with those of logistic model, BP neural network model, and single support vector machine. Experiments show that the prediction level of PCA-SVM model is the best, and the accuracy of the second model is not as good as this model. The error rates of the second type are 14.81%, 14.54%, and 7.67%, respectively, which are higher than those of PCA-SVM model. After comparing the models in this paper, it is found that the components of the model should be extracted first, and then the data operation should be carried out. Such calculation trajectory has a high level of accuracy. The research of this paper provides reference value for solving the financial problems of manufacturing enterprises.

1. Introduction

As one of the core pillar industries in China, manufacturing industry plays an indispensable role in enhancing national strength, upgrading and development of new technologies, and improving people's living standards [1]. In the current competitive market environment, due to poor management, insufficient experience, and technology, more and more manufacturing listed companies began to break out of financial crisis. For example, the capital chain is broken for a long time, the debt cannot be paid off, and the debt crisis is serious. [2]. Local protectionism and trade conservatism also make manufacturing listed companies difficult to develop in the international environment [3]. In order to minimize the possibility of financial crisis of listed manufacturing companies, it is particularly important to carry out financial crisis early warning through data mining and prediction,

and the types of data prediction models built on this basis are also increasing [4].

The experiment adopts a method based on PCA and combined with SVM to construct PCA-SVM model and predict the data. PCA will use matrix transformation to process the values to obtain new eigenvalues and corresponding eigenvectors, then analyze the correlation between the data, leaving only the data with small correlation, and finally analyze the data and obtain the principal components. After obtaining the principal component, SVM will train and learn the sample data based on the principal component and then use it to test the sample data.

The innovation of this paper is that the PCA-SVM model is used to construct multiple financial analysis trajectories, and the results are compared with those of logistic model, BP neural network model, and single support vector machine. In order to fundamentally alleviate and solve the financial

problems of enterprises, this paper no longer uses a single data prediction method, but uses PCA to extract and analyze the principal components. PCA and SVM are organically combined, and SVM is used for operation according to the results of principal component extraction and analysis. Compared with other methods, especially a single SVM model, we can more intuitively see the accuracy of PCA-SVM and get more objective experimental conclusions from the analysis of absolute and relative values.

The structure and content of this article are described and analyzed as follows: the first part mainly expounds the experimental research background and significance of the article in detail, describes the research methods and innovations in more detail, and introduces the structure of the article. The second part mainly describes the selection of principal components, data indicators, and sample data training and testing of PCA-SVM model in detail. The third part mainly explores the experimental results based on the predicted data and actual data, and analyzes the experimental results. The fourth part is the experimental conclusion and puts forward the shortcomings of the research and the prospect of future research.

2. Related Work

The analysis and discussion of finance have always been a hot topic, especially the data prediction related to the financial crisis. Sun et al. [5] artificially combine dynamic financial distress prediction with time weight, using a time-weighted SVM method, using dual expert voting integration, and using error-based expert voting and time-based expert voting and external combination. The experimental results of sample test show that this method is more accurate for dynamic financial distress prediction when time changes [5]. Simian et al. [6] scholars use the hybrid method of multi-core support vector regression, compare and analyze it with the single core support vector regression method, and use gene expression programming and extreme learning machine to fit and predict the sample data. The results show that the multi-core support vector regression method has higher accuracy and efficiency [6]. Zhao et al. [7] used the least squares support vector machine and compared the grid search and genetic algorithm on the basis of PCA, self-excitation threshold autoregression, and other methods, so as to optimize the ant colony of the least squares. The experimental results show that the improved least squares support vector machine is significantly better than all kinds of traditional data prediction, and the accuracy is significantly improved [7]. Kalaiselvi et al. [8] scholars believe that it can accurately predict the stock price and index. An additional weighted value update is added to adjust the balance point between the reverse learning algorithm and BP neural network. The experimental results show that the accuracy and operation efficiency of the improved method have been improved [8]. Munoz Izquierdo et al. combined accounting and audit data and developed the logit prediction model to predict based on the proposed combination of various data indicators. [9].

Taking several bankrupt companies and an equal number of companies with good operating conditions as samples, including other parts, the research makes a horizontal prediction analysis of the research model through different operating characteristics. The research curve tested a different sensitivity and area, and tested its functional ability. The results showed that the prediction and diagnosis accuracy of the new model was high [10]. Gunawardana [11] developed a new neural network to analyze the panel regression considering the five-year cross-sectional data of the selected company sample, which can determine the significant relationship between the leverage ratio, P/E ratio, and the prediction of the company's financial distress. The results show that the prediction accuracy of the model is high [11]. Tunio's et al. [12] studied different kinds of classifiers and used them to fit and classify the sample data. The final results show that the adaptive improved classifier can effectively classify the data and reduce the risk of financial crisis [12]. In order to predict the price of financial market, Henrique et al. [13] used a variety of machine learning models and bibliographic survey technology to analyze the consistency of sample data of different indicators, and finally came to the conclusion that SVM has better effect (Henrique b m et al. 2019) [13]. Thi Vu et al. [14] believe that the prediction of the future financial situation of stock exchange trading companies adopts the combination of SVM and autoregressive method, and takes factor analysis and F-score analysis as the benchmark. The accuracy of the final prediction results is high [14]. The original financial policy and capital allocation behavior of enterprises are based on economic prosperity, financial health, inflation, and excess liquidity. The financial crisis has reversed the financial management environment, and enterprises must adjust the original financial policy and capital allocation structure. The financial policy should be changed from active to passive defense, and the capital allocation structure should also be adjusted according to the new financial policy, Reference [15].

It can be seen from the above research that PCA and SVM have been widely used in financial crisis data prediction, but the research on the combination of PCA and SVM is less. The research combines the methods of PCA and SVM to form the PCA-SVM model and compares it with the other three prediction models. It is expected that the model studied can make an accurate early warning for the financial crisis.

3. Model Building

3.1. Basic Construction Methods of PCA and SVM Models. As a statistical analysis method used for many years, the basic principle of PCA is that when there are too many variables, it will mine and find the correlation between variables and the subject, delete redundant related variables, and create new variables as few as possible according to the overlap between variables, so as to make the correlation between these variables as low as possible or no correlation, and minimize the number of variables without affecting the accuracy of the main research results and reducing the loss of information [16].

The new variable obtained by PCA is called principal component. When the observed value is m and the attribute value is p, the number of principal components is $\min(m-1, p)$. If there are n row data and a characteristic dimensions in the research object, x_{ij} is the i dimension attribute in the j row data, and there is matrix X at this time. Matrix X is a $n \times a$ matrix, and its covariance matrix is shown as

$$C = \frac{1}{n}X^{t+1}. \tag{1}$$

In (1), t is the number of time series. The covariance matrix C is a symmetric matrix of $a \times a$, and its diagonal is the variance of each eigenvalue. Covariance matrix C is also a real symmetric matrix, which has some properties of real symmetric matrix, and these properties can be used to obtain several nonzero eigenvectors to form a new matrix. From matrix E, a new matrix Λ can be obtained, as follows:

$$\Lambda = E^{t+1}C. \tag{2}$$

In order to reduce the redundant data, the data in the characteristic matrix X are converted to another characteristic space to obtain a new matrix Y. The eigenvector e_i corresponding to each eigenvalue needs to satisfy equation (3). In equation (3), $i = 1, 2, ..., p$ and e_{ij} are j vectors of e_i

$$\sum_{j=1}^{p} e_{ij}^2 = 1. \tag{3}$$

Since each characteristic of matrix Y is expected to be linear independent, its covariance matrix D is also a diagonal matrix, its diagonal is variance, that is, the blank is covariance. When the covariance is 0, it means that the two vectors are orthogonal. The eigenvectors of the matrix C form a new matrix Z, and the covariance matrix D is shown as

$$D = Z^{t+1}C = \frac{1}{n}Y^{t+1}. \tag{4}$$

Each value on the diagonal of matrix D is the eigenvalue of covariance matrix C [17]. The eigenvalues and eigenvectors in the matrix D are sorted from small to large and from left to right, respectively, and then the first k are taken for compression conversion to obtain the dimension reduced data matrix Y. In matrix X, there is no correlation between the principal components of the linear combination of its attributes, and the principal components have the maximum variance in the linear combination of various attributes. The proportion of total variance from the first to n the second principal component decreases in turn. In general, several principal components with a cumulative proportion of more than 85% can be considered as less information loss, which can be adopted [18]. The flowchart of the whole process is shown in Figure 1.

Use the hinge loss function to calculate the empirical risk. The advantage of SVM can balance the search for complex models and learning ability even without sufficient large sample data [19]. The classification of SVM usually uses sigmoid function, as shown in (5), which is the mapping of sigmoid function.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^t}}. \tag{5}$$

The three key points of SVM are classification interval, duality, and kernel. These three key points determine the practicability of SVM, the principle of minimizing structural risk and the separation of dimension space. In Reference [20], two different samples can be distinguished by a linear function. At this time, assuming that the initial training set can be linearly divided by the hyperplane, the interval between two different categories of samples should be the largest when the optimal hyperplane is established. The classification principle is shown in Figure 2.

When the samples are classified, if these data are nonlinear, they will be selected to solve it. At the same time, the dual method is used to reduce the complexity of spatial dimension upgrading and make it simpler. In the process of dimensionality reduction, we will reduce the number of features, which requires us to delete the data. However, less data will lead to less information available to the model, and the performance of the model may be affected. Find a way to help us measure the amount of information that features carry. In the process of dimensionality reduction, we can not only reduce the number of features, but also retain most of the effective information. Merge those features with duplicate information, delete those features with invalid information, and so on. Gradually create a new feature matrix that can represent most of the information of the original feature matrix with fewer features. The dual characteristic makes it possible to obtain the corresponding classification function only by the operation of sample inner product and operate it. The principle of obtaining the optimal hyperplane is shown in Figure 3.

Based on the idea of kernel method, the operation of SVM can map the data difficult to be separated by plane into high-dimensional space through the method of feature transformation to obtain the classification plane, Reference [21]. The dot product operations $\lambda(x_i)$ and $\lambda(x_j)$ replace the operation with the low-dimensional space, and obtain equation (6) and the corresponding decision function equation (7) through dual transformation.

$$\max W(\beta) = \sum_{i=1}^{n} \beta_i - \frac{1}{2} \sum_{i,j=1}^{n} \beta_i \beta_j y_i y_j K(x_i, x_j), \tag{6}$$

$$f(x) = sign\left\{ \sum_{i=1}^{n} \beta_i y_i K(x_i, x) + g \right\}. \tag{7}$$

In equations (4) and (5), $i = 1, 2, \ldots, n$. The meaning of β in formula (6) and formula (7) is shown as follows:

$$\sum_{i=1}^{n} \beta_i y_i = 0, 0 \le \beta \le C. \tag{8}$$

In order to obtain the maximum classification interval, the decision function and β are transformed into a new objective function Ψ, as shown in
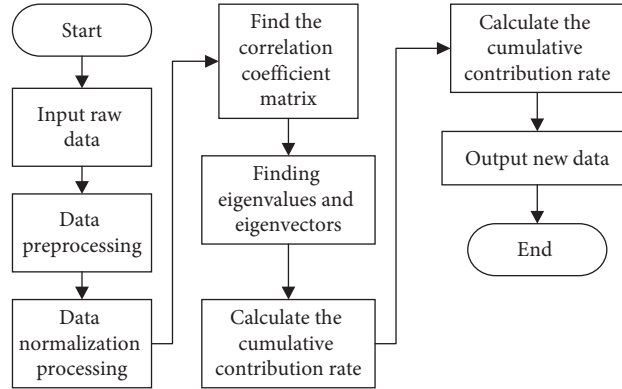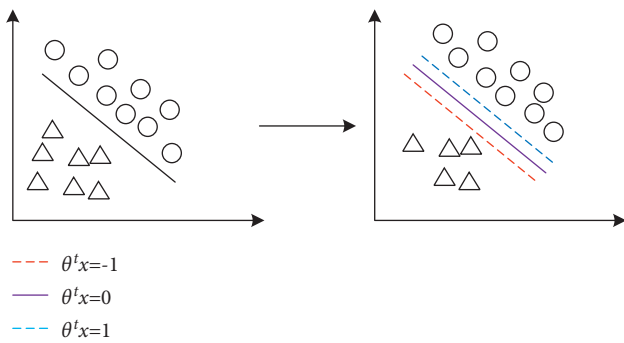
FIGURE 1: Simple flowchart of PCA.

$\theta^t x = -1$
$\theta^t x = 0$
$\theta^t x = 1$

FIGURE 2: Classification principle of SVM.
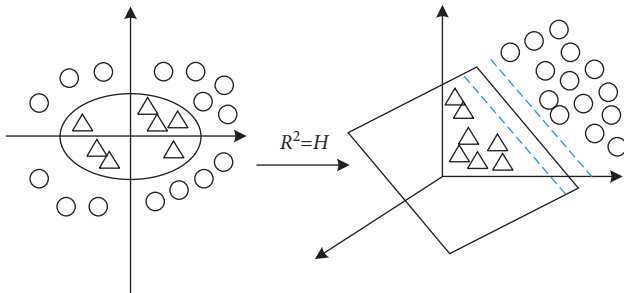

$R^2 = H$

FIGURE 3: Optimal classification hyperplane.

$$\Psi(\beta) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n}\beta_i. \qquad (9)$$

In order to integrate the respective advantages of PCA and SVM, the experiment will build a PCA-SVM combined prediction model. There are two combination modes. The first is to linearly combine the principal components obtained from the original data through principal component analysis and the results predicted by SVM, calculate the output, and take it as the final result. The whole process is shown as

$$y = b_1 y_1 + b_2 y_2. \qquad (10)$$

Equation (7) is the corresponding combination coefficient. The second method is to input the original data into the SVM model through principal component analysis and get the final result through SVM training and prediction. The expression of this method is shown as

$$y = SVM[PCA(X)]. \qquad (11)$$

Principal component analysis can eliminate the correlation between the evaluation indicators because principal component analysis has formed independent principal components after transforming the original index. The higher the correlation degree between the indicators with practice, the better the effect of principal component analysis. It can reduce the workload of index selection for other evaluation methods. Because it is difficult to eliminate the related impact between the evaluation indicators, it takes a lot of energy to select indicators. Principal component analysis can eliminate this correlation, so it is relatively easy to select indicators. When constructing PCA-SVM financial early warning model in the experiment, we mainly consider the advantages of principal component analysis in dimensionality reduction, while method 1 is not as good as method 2 in giving full play to its advantages, and method 1 needs to consider the weight, which will artificially bring subjective errors, so we use more objective method 2. Therefore, the construction of PCA-SVM model is to extract the principal components to obtain meaningful data, and then use the Gaussian radial basis kernel function to input the principal components into the vector set. The subsequent calculations are based on the principal components.

3.2. Data Processing and Comparative Prediction of PCA-SVM Financial Crisis Early Warning Model. The selection of samples includes two parts. The first part is to select the sample indicators to predict, and the second part is to select the sample data of the financial status of listed manufacturing. The selection of early warning indicators needs to fully include the company's financial stability. Therefore, when evaluating and selecting early warning indicators, we must consider the reflection function, comparison function, evaluation function, and prediction function. In addition, the principles of comprehensiveness, systematicness, and operability need to be followed [22]. A total of 19 indicators were selected in the experiment. Because not all of the 19 financial indicators preliminarily

selected are of great value, and the distribution of the research sample data is unknown, K-S test is used to determine the significance of these 19 indicators. Among the 19 K-S test results, the p value of only 4 indexes is greater than 0.05, which obviously does not obey the normal distribution. For this non-normal distribution, the experiment uses Mann–Whitney U's nonparametric test and spss25 to complete this process [23]. Among the test results obtained, 11 of the 19 primary indexes have a p greater than 0.05, with significant difference, while the remaining 8 data do not. Therefore, the 11 data indicators retained in the experiment are current ratio, quick ratio, asset liability ratio, long-term debt equity ratio, total asset turnover rate, total asset return rate, operating profit rate, cost utilization rate, and capital preservation and appreciation rate, so as to ensure the comprehensiveness and objectivity of the index system.

For the sample data of the financial situation of listed manufacturing companies, the object of the experiment is listed manufacturing companies in China. The companies studied belong to Shanghai and Shenzhen A shares. The negative samples are ST and ∗ ST Companies in listed manufacturing companies. The positive sample is non-ST and non- ∗ ST Companies with the corresponding asset scale. The PCA-SVM is trained through the training sample group, and the prediction ability of PCA-SVM model is tested through the test sample group [24].

This study selects manufacturing companies that have been in ST or ∗ ST status in the past three years and companies that have not been in ST or ∗ ST status as samples from the A share main board markets in Shanghai and Shenzhen. In order to train and predict the model, the experiment needs to divide the company sample into two parts, with a ratio of 4 : 1, 80% of the samples as the training samples and the rest as the test samples. In order to make the limited sample data play a more effective role, the experiment randomly divided five samples into five data sets for training and testing. In this way, five support vector machine models can be obtained and tested, respectively, so as to improve the efficiency. All data need to be processed before use. The experiment adopts the commonly used Z-score method for preprocessing, as shown in (12). In (12), z_{ij} is the sample value after standardized treatment, and x_{ij} represents the j index value in the i company.

$$z_{ij} = \frac{x_{ij} - \overline{x}_j}{\sqrt{\text{var}(x_j)}}. \tag{12}$$

After obtaining the principal component, the maximum and minimum specification method needs to be used for the original data, as follows:

$$g' = \frac{g - \min a}{\max a - \min a}. \tag{13}$$

Kmo test and Bartlett sphere test must be carried out first. The significance level in the test results is 0, which means that there is a linear correlation between these data, which is suitable for principal component analysis. Since the sum of variance percentage and cumulative is greater than 85%, it indicates these indicators and can be used. On this basis, SPSS is used to analyze the principal components to obtain the corresponding principal component score coefficient matrix, and the obtained results and actual results are used to analyze the accuracy of prediction [25].

The first mock exam is to compare the prediction effect of PCA-SVM on the financial crisis. Three models are selected for comparison, namely, logistic model, BP neural network model, and SVM model. BP neural network model also has strong learning ability, and a single SVM can be used to compare with PCA-SVM. For logistic model, when the significance P is greater than 0.05, it is suitable to use this model for fitting. The probability construction of the model is shown in (14). In (14), b is the score model, f is all extracted independent variables, and P is the probability. When P is greater than 0.5, it is considered as a crisis enterprise and gives an early warning. The grouping method of logistic model is the same.

$$\ln \frac{P}{1 - P} = b_1 f_1 + b_2 f_2 + \dots + b_n f_n. \tag{14}$$

When building the BP neural network model, this paper needs to further include the structure and content required by the research. The number of nodes is shown in formula (15). a is the adjustment constant, and 8 is selected in the experiment. Therefore, the number of nodes is 11, and the model is a simple structure. Grouping and the selection of training set and test set also follow the grouping method of PCA-SVM model.

$$h = \sqrt{n + u + a}. \tag{15}$$

When using a single SVM model, we only need to repeat the research steps without principal component extraction, then write the SVM program, train through the training samples, and then get the prediction results through the test samples. The grouping method is the same.

4. Result Analysis

4.1. PCA-SVM Principal Component Extraction Results and Significance Analysis. After SPSS analysis, the score coefficients corresponding to the principal components are obtained and classified in turn. The score coefficient corresponding to the principal component $F1$ is shown in Figure 4. At this time, it is found that the score coefficients of the five indicators of return on assets, net profit rate of total assets, net income rate, expense profit rate, and operating profit rate are significantly greater than those of other indicators. Therefore, it can be concluded that the principal component $F1$ is mainly determined by the return on assets, net profit rate of total assets, net income rate, expense profit rate, and operating profit rate. It can be judged that $F1$ represents the profitability of the company.

The score coefficient corresponding to the principal component $F2$ is shown in Figure 5. The asset liability ratio, long-term debt equity ratio, and capital preservation and appreciation rate in the score coefficient are significantly greater than those of other data. Therefore, it can be concluded that the main data composition of $F2$ is asset liability
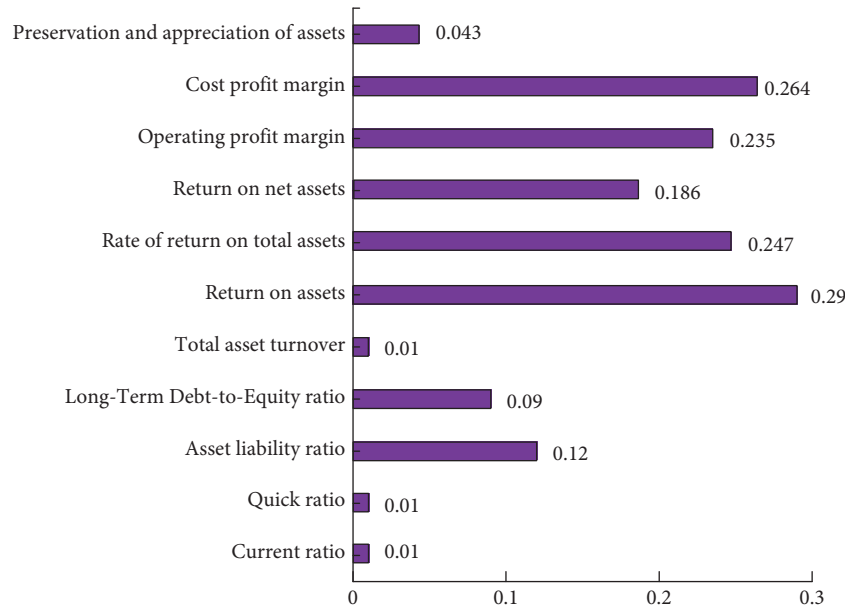
FIGURE 4: Score coefficients of principal component $F1$.
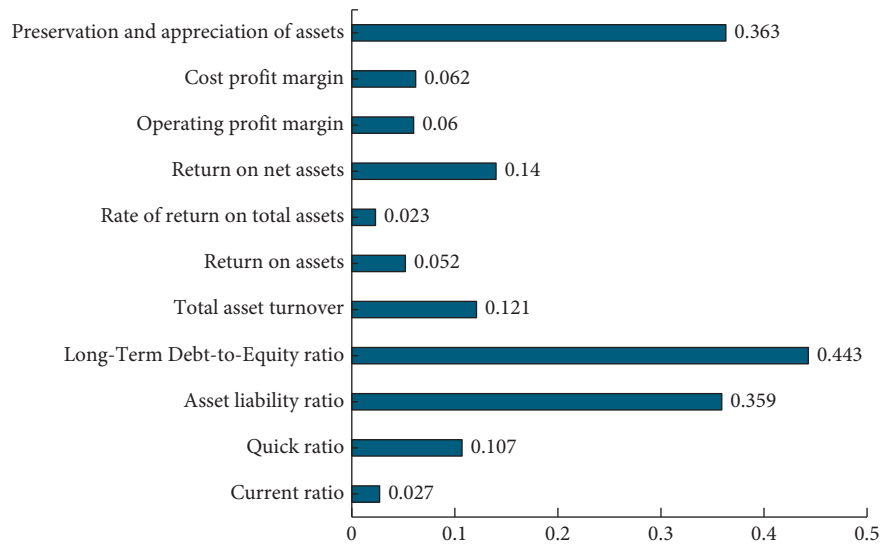
FIGURE 5: Score coefficients of principal component $F2$.

ratio, long-term debt equity ratio, and capital preservation and appreciation rate, indicating that $F2$ represents the company's debt repayment ability and capital preservation ability.

The score coefficient corresponding to the principal component $F3$ is shown in Figure 6. The current ratio and quick ratio are significantly greater than the score coefficients of other indicators, indicating that the main constituent indicators of $F3$ are current ratio and quick ratio, which means that $F3$ measures the short-term debt repayment ability of the formula.

The score coefficient corresponding to the principal component $F4$ is shown in Figure 7. The score coefficients are significantly higher than other data. It can be seen that $F4$ is the main component to measure the liquidity of the company's assets and the profitability of its own capital.

4.2. Analysis of Prediction Results of PCA-SVM and Comparative Model. The finally recorded data indicators are the average accuracy, average class I error rate, and average class II error rate of 5 groups of prediction results of each method. The class I error refers to the fact that there
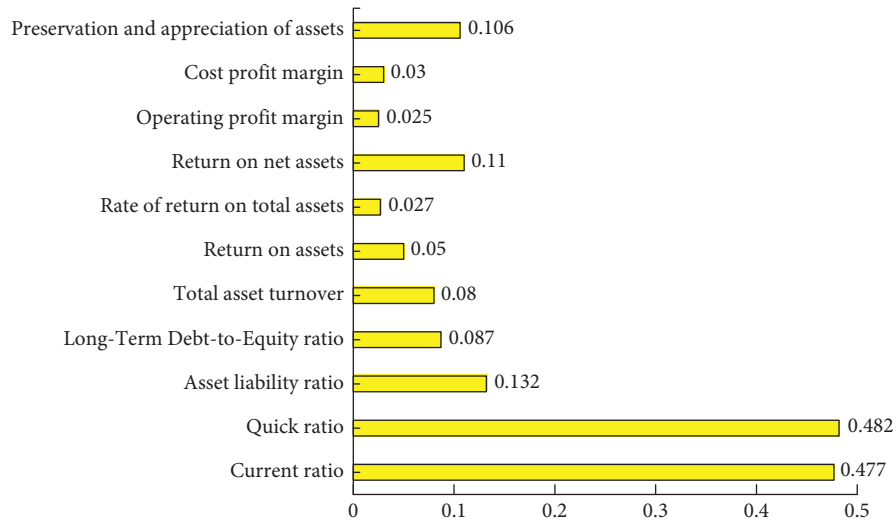
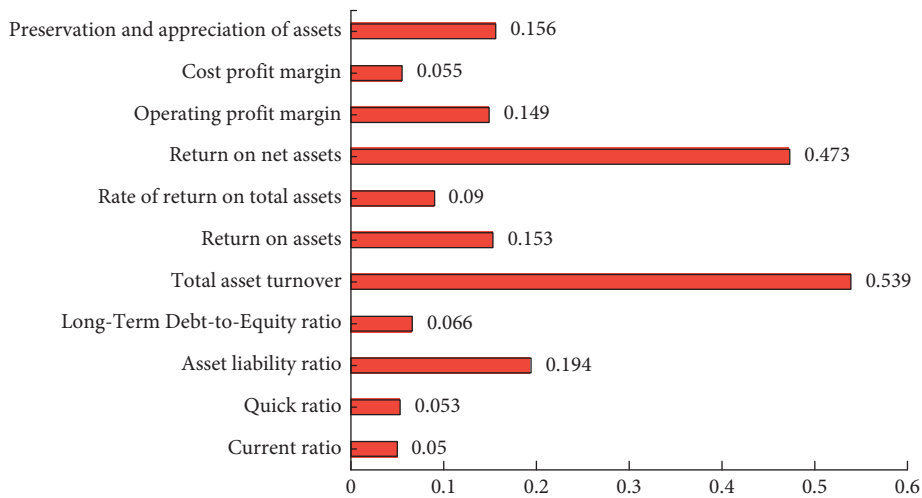FIGURE 6: Score coefficients of principal component $F3$.

FIGURE 7: Score coefficients of principal component $F4$.

is no financial crisis but is predicted to be an outbreak of financial crisis. Obviously, the error rate of the second type of error is more important than that of the first type. The average accuracy of each method is shown in Figure 8.

The average type I error rate and type II error rate of each method are shown in Figure 9 The average class I error rate and the average class II error rate of PCA-SVM are significantly lower than those of the other three methods; especially, the class II error rate is only 3.64%, so the prediction result of PCA-SVM is reliable.

The calculation accuracy results of different methods are shown in the figure. The comparison results of the first type of errors have different accuracy indication results compared with the comparison results of the second type of errors. It can be seen from Figure 10 that the calculation accuracy of each of the four methods is very high. Because the fault
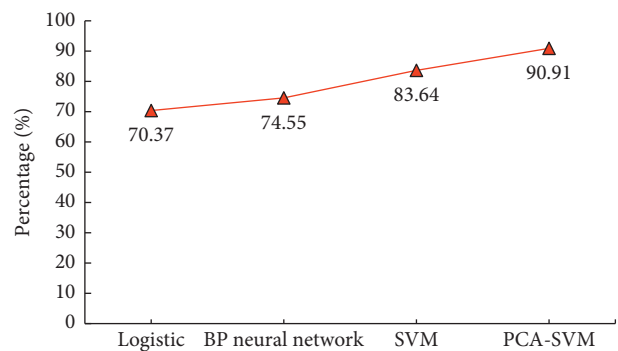

FIGURE 8: Average accuracy of different methods.

tolerance rate of the results of the first and second types is very low, these contents can be used as the background content of the early warning model.
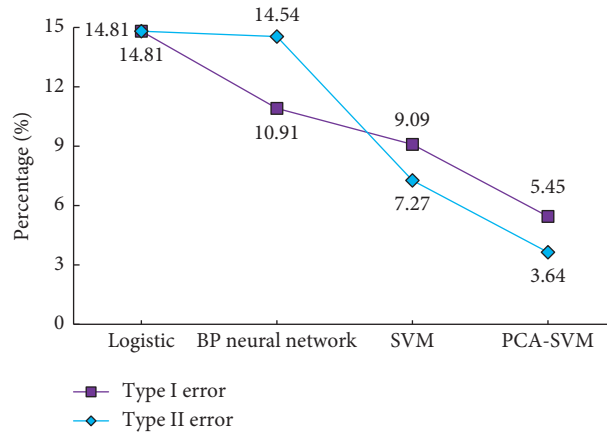
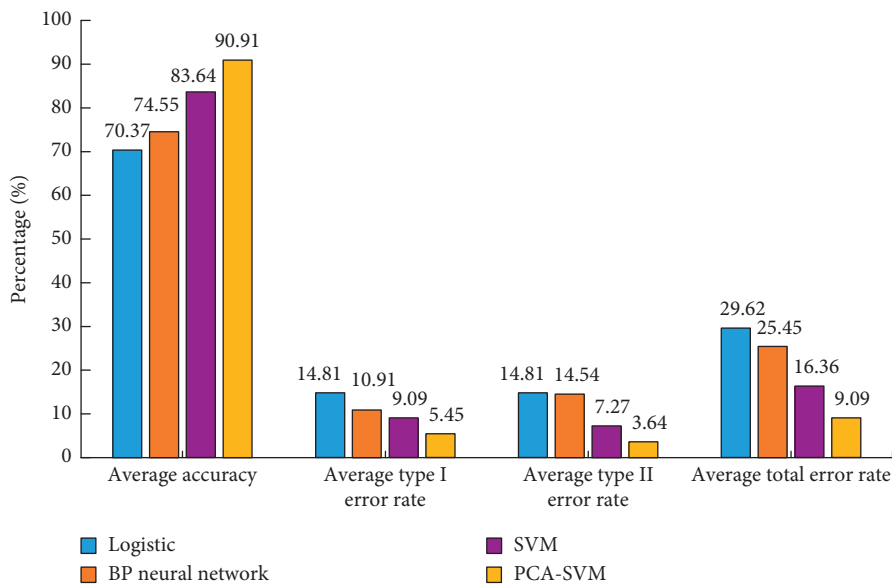FIGURE 9: Average type I error rate and of type II error rate different methods.

FIGURE 10: Comparison of average accuracy between different methods.

5. Conclusion

The model proposed in this paper has higher accuracy in different crisis estimates, which is significantly higher than logistic model, BP neural network model, and single SVM model. PCA-SVM is not only better than traditional methods, but also better than a single SVM model. The results show that the combination of support vector machine and principal component analysis can play an auxiliary role and can be applied to manufacturing industry. This paper no longer uses a single data prediction method, but uses principal component analysis to extract and analyze the principal components. The principal component analysis and support vector machine are organically combined. According to the results of principal component extraction and analysis, support vector machine is used for operation. Compared with other methods, especially a single SVM model, we can more intuitively see the accuracy of PCA-SVM and get more objective experimental conclusions through the analysis of absolute and relative values.

Although the research has made some achievements, the number of research samples is relatively small, lacking universality and representativeness. In the follow-up study, more and more representative sample data need to be selected, which also needs to be improved in further research in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

finance effectively supports research teams for high-quality development of the real economy (No. 2021XJTD02); university-level scientific research fund project of Xi'an Eurasia University: Research on operation mode of supply chain finance ecosphere in Shaanxi (Xi 'an) Pilot Free Trade Zone under "double circulation" background (No. 2021XJSK28); 2021 "New Liberal Arts" research and Reform Practice project of Xi'an Eurasia University: Research on the mode of production-university-research cooperation education under the background of "New Liberal Arts" (No. 2021WKXJ001).

References

[1] Y. Gong, J. Zhu, Y. Chen, and W. D. Cook, "DEA as a tool for auditing: application to Chinese manufacturing industry with parallel network structures," *Annals of Operations Research*, vol. 263, no. 1-2, pp. 1–23, 2018.

[2] A. Ehlert, "The socio-economic determinants of COVID-19: a spatial analysis of German county level data," *Socio-Economic Planning Sciences*, vol. 78, Article ID 101083, 2021.

[3] N. Khidil, M. A. S. Zaidi, and Z. Abdul Karim, "Fragmentation of international production and business cycle synchronization: new evidence pre and during global financial crises," *Sustainability*, vol. 13, no. 8, p. 4131, 2021.

[4] O. JI, O. N, P. O, and A. SN, "Evaluation of Financial Soundness of Listed Commercial Banks in India: An Application of Bankometer S-Score Model," *Journal of Finance and Marketing*, vol. 02, 2018.

[5] J. Sun, H. Fujita, Y. Zheng, and W. Ai, "Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods," *Information Sciences*, vol. 559, pp. 153–170, 2021.

[6] D. Simian, F. Stoica, and A. Brbulescu, "Automatic optimized support vector regression for financial data prediction," *Neural Computing & Applications*, vol. 32, no. 1, pp. 1–14, 2020.

[7] D. Zhao, J. Ding, and S. Chai, "Systemic financial risk prediction using least squares support vector machines," *Modern Physics Letters B*, vol. 32, no. 17, Article ID 1850183, 2018.

[8] K. Kalaiselvi, K. Velusamy, and C. Gomathi, "Financial prediction using back propagation neural networks with opposition based learning," *Journal of Physics: Conference Series*, vol. 1142, no. 1, Article ID 012008, 2018.

[9] N. Munoz-Izquierdo, E. K. Laitinen, M. Camacho-Minano, and D. Pascual-Ezama, "Does audit report information improve financial distress prediction over Altman's traditional Z-Score model?" *Journal of International Financial Management & Accounting*, vol. 31, no. 1, pp. 65–97, 2020.

[10] S. H. Shetty and T. N. Vincent, "The role of board independence and ownership structure in improving the efficacy of corporate financial distress prediction model: evidence from India," *Journal of Risk and Financial Management*, vol. 14, no. 7, p. 333, 2021.

[11] K. D. Gunawardana, "An analysis of financial distress prediction of selected listed companies in colombo stock exchange," *International Journal of Sociotechnology and Knowledge Development*, vol. 13, no. 2, pp. 48–70, 2021.

[12] F. Tunio, Y. Ding, A. Nabi, and A. Kinza, "Financial distress prediction using adaboost and bagging in Pakistan stock exchange," *Journal of Asian Finance Economics and Business*, vol. 8, no. 1, pp. 665–673, 2021.

[13] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Literature review: machine learning techniques applied to financial market prediction," *Expert Systems with Applications*, vol. 124, no. 6, pp. 226–251, 2019.

[14] L. Thi Vu, L. Thi Vu, N. Thu Nguyen, P. Thi Thuy Do, and D. Phuong Dao, "Feature selection methods and sampling techniques to financial distress prediction for Vietnamese listed companies," *Investment Management and Financial Innovations*, vol. 16, no. 1, pp. 276–290, 2019.

[15] A. Kamaluddin, N. Ishak, and N. F. Mohammed, "Financial distress prediction through cash flow ratios analysis," *International Journal of Financial Research*, vol. 10, no. 3, p. 63, 2019.

[16] M. Sun, Y. Wang, and J. Liu, "Generalized Peaceman-Rachford splitting method for multiple-block separable convex programming with applications to robust PCA," *Calcolo*, vol. 54, no. 1, pp. 77–94, 2017.

[17] K. Batpurev, S. J. Sinclair, O. Avirmed, M. P. Scroggie, K. Olson, and M. D White, "Stakeholders from diverse backgrounds make similar judgments about ecological condition and collapse in Mongolian rangelands," *Conservation Science and Practice*, vol. 4, no. 1, p. e574, 2022.

[18] A. R. M. A. Basher and S. J. Hallam, "Aggregating statistically correlated metabolic pathways into groups to improve prediction performance," in *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies*, pp. 49–61, Singapore, January2022.

[19] Q. Ding, Z. Shao, X. Huang, O. Altan, and Y Fan, "Improving urban land cover mapping with the fusion of optical and SAR data based on feature selection strategy," *Photogrammetric Engineering & Remote Sensing*, vol. 88, no. 1, pp. 17–28, 2022.

[20] B. Ibrahim, S. Suppiah, N. Ibrahim et al., "Diagnostic power of resting-state fMRI for detection of network connectivity in Alzheimer's disease and mild cognitive impairment: a systematic review," *Human Brain Mapping*, vol. 42, no. 9, pp. 2941–2968, 2021.

[21] A. Tyagi, S. Tyagi, A. Agrawal et al., "Early warning scores at time of ICU admission to predict mortality in critically ill COVID-19 patients," *Disaster Medicine and Public Health Preparedness*, vol. 3, no. 2, pp. 1–5, 2021.

[22] G. F. Shen and Y. W. Zhang, "Research on financial early-warning based on GIHS improved BP_AdaBoost algorithm," *MATEC Web of Conferences*, vol. 173, no. 1, Article ID 03004, 2018.

[23] J. Dedecker and G. Saulière, "The mw," *Mathematical Methods of Statistics*, vol. 26, no. 2, pp. 111–133, 2017.

[24] S. Bakheet, "An SVM framework for malignant melanoma detection based on optimized HOG features," *Computation*, vol. 5, no. 4, p. 4, 2017.

[25] Y. Liu, L. Chen, and C. Zhu, "Improved robust tensor principal component analysis via low-rank core matrix," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1378–1389, 2018.